Magazine
R124

Correspondence

Sequence similarity in structurally dissimilar proteins

Stefan Tomiuk and Kay Hofmann

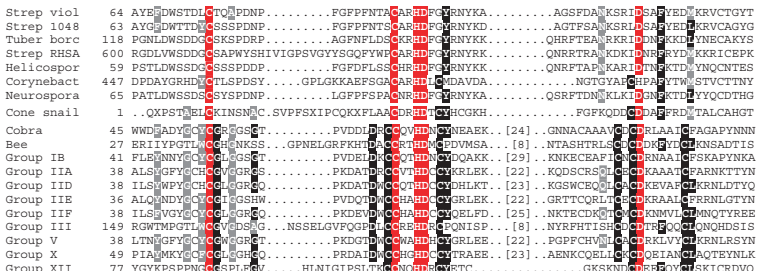
It is a central tenet of protein evolution that the three-dimensional structure of a protein family is better conserved than the sequences themselves. So structural similarities between related proteins should be detectable over longer evolutionary distances and be most useful for the functional prediction of proteins without closely related homologs. Comparative sequence and structural analysis of prokaryotic phospholipase A2 (PLA2) from *Streptomyces violaceoruber* reveals a clear violation of this dogma. While the enzymatic properties of this enzyme match closely those of eukaryotic secretory PLA2s, the corresponding structures appear very different, and no appreciable structural similarity is detected by a number of comparison methods. Nevertheless, contemporary sequence analysis methods demonstrate a highly significant relationship between the sequences of both families and provide a structurally correct alignment of the active site residues.

The recent literature documents many cases of unexpected structural conservation in the complete absence of detectable sequence relatedness. Although in many of those cases the purported lack of sequence similarity does not stand closer scrutiny with sophisticated sequence comparison methods, there clearly are some impressive examples of structure-based functional predictions where sequence comparisons fail [1]. Here, we describe an unusual case where this situation appears to be reversed: related proteins with an unambiguously detectable sequence similarity but whose structures have diverged beyond recognition.

Secretory (s)PLA2s are a class of small enzymes that hydrolyze the 2-acyl ester bond of 1,2-diacylglycero-3-phospholipids [2]. Most eukaryotic cells produce multiple sPLA2 isoenzymes, and several poisonous animals, including bees and snakes, use members of this enzyme class as a major component of their toxins. All secreted PLA2 forms contain Ca(II) ions, which are directly involved in the catalytic reaction [2]. Recently, the first prokaryotic PLA2 has been isolated and cloned from the bacterium *S. violaceoruber*, and subsequently was structurally characterized [3,4]. The enzymatic properties of bacterial PLA2 resemble closely that of the eukaryotic sPLA2 forms, including the strict requirement for Ca(II). But the lack of visible sequence similarity and the fundamentally different structural fold have been interpreted as indicating a distinct evolutionary origin [3].

As indicated in the original report on cloning *Streptomyces* PLA2, sequence database searches with standard methods like BLAST [5] reveal significant similarities to a number of uncharacterized proteins from other bacteria, but fail to show a relationship to established PLA2 forms or to other hydrolytic enzymes. To search for more distant sequence relatives, we constructed generalized profiles [6] from a multiple alignment of the bacterial PLA2 and its reliable BLAST matches (Figure 1, upper part).

Unexpectedly, the result of the profile search clearly demonstrates a significant relationship to a number of established PLA2s: The best match (p<0.01) was the conodipine-M α chain from cone snails, a well-characterized PLA2 toxin [7]. Among the next seven high-scoring sequences, six corresponded to known eukaryotic PLA2s, including the group XIV enzyme from *Drosophila* and the mammalian group XIII enzymes. The seventh sequence was an uncharacterized *Pseudomonas* protein, another likely PLA2. The best-scoring non-PLA2 was the MAP kinase ERK4, which reached an insignificant p value of only 0.6. As expected for profile searches, the significance values of other eukaryotic PLA2 sequences further improved after incorporating conodipine-M in a subsequent cycle of iterative profile refinement. As shown in Figure 1, two classes of residues are nearly invariant between prokaryotic and eukaryotic PLA2 isoenzymes: two cysteine residues that form a structurally important disulfide bridge, and the polar residues required for catalysis and the coordination of one Ca(II) ion. The eukaryotic PLA2 enzymes typically



Current Biology

Figure 1. Alignment of prokaryotic and eukaryotic sPLA2 families.
Multiple alignment of the prokaryotic PLA2 family and representative members of the eukaryotic sPLA2 families for comparison. Sequence names are in the leftmost column: PLA2 from *S. violaceoruber* and *S. coelicolor*; SCO01048 from *S. coelicolor*; Pst1 from *Tuber borchii*; RHSA from *S. clavuligerus*; p15 from *Helicosporium* sp; CGL2546 from *Corynebacterium glutamicum*; ORF from *Neurospora crassa*; and PLA2s from cone snail, cobra, bee, and various human isoenzymes. Only the most conserved part of the catalytic domain is shown. The position of the fragment relative to the complete sequence is indicated by numbers. Residues invariant or conservatively replaced in at least 50% of the sequences are printed on black or grey background, respectively. The catalytically important residues conserved between all PLA2 subtypes are printed on red background.

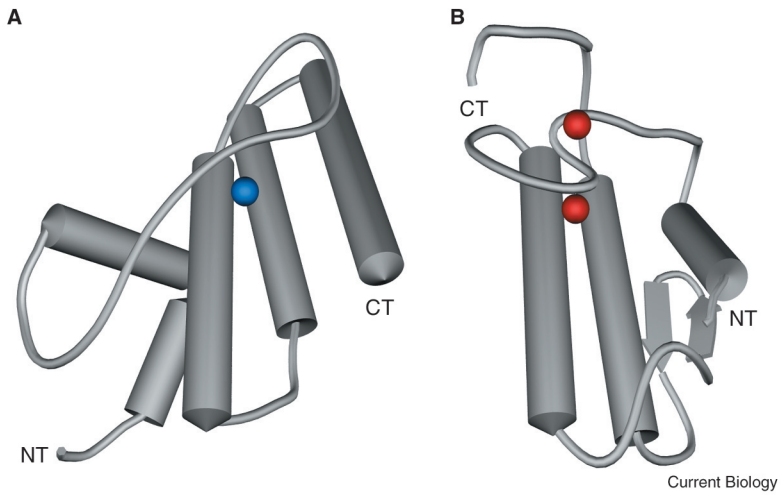


Figure 2. Structural comparison between prokaryotic and eukaryotic sPLA2.
Side-by-side view of structurally aligned PLA2 proteins. (A) Prokaryotic PLA2 from *S. violaceoruber* (pdb:1KP4), with a single coordinated Ca(II) ion. (B) Eukaryotic sPLA2 from cobra venom (pdb:1POB), with two coordinated Ca(II) ions. Only the two central helices with the catalytically important residues and one Ca(II) ion show structural correspondence. Structural superposition was performed with SPDBV [11] using the Ca(II)-coordination sphere as anchor residues.

contain a second Ca(II) ion and possess additional disulfide bridges.

The alignment shown in Figure 1 allowed us to calculate a superposition of the PLA2 structures. Figure 2 shows structural cartoons for the *Streptomyces* enzyme (pdb: 1KP4) and the cobra PLA2 (pdb: 1POB), a representative eukaryotic sPLA2 structure. The only regions of the molecules yielding a reasonable superposition are the two long central helices that contain the residues important for catalysis and Ca(II)-coordination. Other parts of the structure, including the amino-terminal half of the disulfide bridge are in a different structural context, giving the overall fold a very dissimilar appearance. More important than the visually perceived structural similarity — or the lack thereof — is the ability of structural comparison algorithms to detect a statistically significant relationship between the structures. Both structural comparison programs used, DALI [8] and VAST [1], failed on this account. The top-scoring DALI matches are unrelated to PLAs, but are structurally more similar. The first genuine eukaryotic PLA2 structure is found at rank 190 with a Z-score of 2.0, far below any sensible significance threshold. Similarly, VAST did not place any PLA2 among the top-scoring matches.

From comparing the structures in Figure 2, it becomes clear that the two enzyme classes are true homologs: the sequences as well as the structures are distantly related. It is the relative degree of divergence in structure and sequence that sets this case apart from those discussed previously [9], where a limited structural correspondence usually goes along with an even lower degree of sequence conservation. In the case of prokaryotic PLA2, the structural conservation is confined to a region too small to yield significant comparison scores, at least in the context of the non-conserved portion of the structure. The structural comparison clearly performs better when using only the conserved central region, although the resulting DALI-scores are still not significant. Moreover, in a structural genomics setting, this information is typically not available beforehand.

Apparently, modern profile-based sequence comparison methods have advantages in handling such extreme cases of 'focal conservation'. It is to be expected that structural comparison methods too could profit from the 'profile approach'. A comparison of relatively closely related structures would reveal which positions of the fold are less variable than others. Those structurally conserved elements could then be assigned a higher weight in subsequent comparison cycles, resulting in an iterative refinement process as it is routinely used in profile and HMM methods [10].

References

1. Gibrat, J.F., Madej, T., and Bryant, S.H. (1996). Surprising similarities in structure comparison. Curr. Opin. Struct. Biol. *6*, 377-385.
2. Valentin, E., and Lambeau, G. (2000). Increasing molecular diversity of secreted phospholipases A(2) and their receptors and binding proteins. Biochim. Biophys. Acta *1488*, 59-70.
3. Sugiyama, M., Ohtani, K., Izuhara, M., Koike, T., Suzuki, K., Imamura, S., and Misaki, H. (2002). A novel prokaryotic phospholipase A2. Characterization, gene cloning, and solution structure. J. Biol. Chem. *277*, 20051-20058.
4. Matoba, Y., Katsube, Y., and Sugiyama, M. (2002). The crystal structure of prokaryotic phospholipase A2. J. Biol. Chem. *277*, 20059-20069.
5. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389-3402.
6. Bucher, P., Karplus, K., Moeri, N., and Hofmann, K. (1996). A flexible motif search technique based on generalized profiles. Comput. Chem. *20*, 3-23.
7. McIntosh, J.M., Ghomashchi, F., Gelb, M.H., Dooley, D.J., Stoehr, S.J., Giordani, A.B., Naisbitt, S.R., and Olivera, B.M. (1995). Conodipine-M, a novel phospholipase A2 isolated from the venom of the marine snail Conus magus. J. Biol. Chem. *270*, 3518-3526.
8. Holm, L., and Sander, C. (1995). Dali: a network tool for protein structure comparison. Trends Biochem. Sci., 20: 478-480.
9. Holm, L., and Sander, C. (1997). New structure — novel fold? Structure *5*, 165-171.
10. Hofmann K. (2000). Sensitive protein comparisons with profiles and hidden Markov models. Brief. Bioinform. *1*, 167-178.
11. Guex N., Peitsch M.C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis *18*, 2714-2723.

Bioinformatics Group, MEMOREC Stoffel GmbH, Stöckheimer Weg 1, D-50829 Köln, Germany.
E-mail: kay.hofmann@memorec.com